

# **Quantitative Lexicology and Variational Linguistics (QLVL) and the study of Dutch: a goal as well as a tool**

Jocelyne Daems   Karlien Franco   Laura Rosseel   Thomas Wielfaert  
Quantitative Lexicology and Variational Linguistics (QLVL), KU Leuven  
{jocelyne.daems; karlien.franco; laura.rosseel; thomas.wielfaert}@arts.kuleuven.be  
<http://wwwling.arts.kuleuven.be/qlvl/>

## **1 The research group QLVL**

The Quantitative Lexicology and Variational Linguistics (QLVL) research group from the Linguistics research unit at the Arts faculty has a strong and long-standing research tradition. Within the framework of Cognitive Linguistics, language variation is approached from a predominantly quantitative and corpus-based perspective. This area of investigation entails two specific interests: lexical variation at large, both from a synchronic and a diachronic point of view, and the interaction between internal (structural) and external (lectal) variation. Extensions in various directions further solidify the research group's methodological, descriptive and theoretical foundations.

## **2 The study of Dutch: different approaches to a pluricentric stratified language**

One of the principle preoccupations of the research group is the study of ongoing changes in Belgian Dutch and its relationship with Netherlandic Dutch. To this end, large text corpora of both written and spoken Dutch (from both national varieties) in various registers (from archived newspapers to new online media) are analysed. This corpus-based approach is occasionally complemented by experimental techniques or digitalised dictionary databases. Additionally, attitudinal research looks into the implicit judgement of language users vis-à-vis variation in language.

### **2.1 Pluricentrism, or having more than one national language variety**

Dutch is considered a pluricentric language (Clyne, 1992), since it has more than one national variety: Netherlandic Dutch, spoken in the Netherlands, and Belgian Dutch, spoken in Flanders, the northern part of Belgium. The accompanying language variation relates, among other things, to the process of linguistic standardization that evolved differently in both regions (Geeraerts, 2003). Alternatives are attested for instance in pronunciation, in word choice and meaning, or in sentence constructions. From a diachronic point of view, language policy and changes in society each show a noteworthy influence on the degree of variation.

### **2.2 Stratification, or having clearly delineated language layers**

One of the consequences of the different linguistic standardisation developments in Flanders and the Netherlands returns in the stratificational spectrum of both languages. Whereas in the Netherlands the border between the long-established standard Netherlandic Dutch and the colloquial variant is rather fuzzy, the more recent construal of standard Belgian Dutch can be clearly demarcated from colloquial Belgian Dutch (CBD), or “tussentaal” (Geeraerts, 2011; Geeraerts and Van de Velde, 2013). Regional and dialectal language use are present in both linguistic regions and colour colloquial and standard language use to varying degrees.

### **3 Methodology: usage-based approaches**

In order to be able to deal with the data quantities and complexities a usage-based approach entails, QLVL's research method almost necessarily relies on a wide range of methodological techniques. Moreover, the variety of research topics and approaches implies the use of relevant techniques for each specific line of research. In addition to the use of traditional and well-established methods, more recently developed techniques are employed (and compared to classic approaches) and new methods are being created to deal with the research situation as aptly as possible.

#### **3.1 On implicit language attitudes and reaction times**

People have explicit, but also unconscious attitudes towards languages and language varieties. To uncover these unconscious attitudes, a number of QLVL researchers have recently developed the auditory affective priming (AAP) method, an innovation of the affective priming paradigm used in social psychology (Speelman et al. , 2013). A classic AP experiment measures the reaction time a respondent needs to categorise a target picture as negative or positive after being presented with a prime picture. If the prime has the same polarity as the target picture, it has been found that the reaction time is shorter than when prime and target are contrastively polarised. Analysing reaction times, then, allows to determine whether a stimulus is experienced positively or negatively. AAP brings innovation to this technique by replacing the visual primes from the original technique by auditory ones which allows linguists to study language attitudes.

#### **3.2 From manually verified to automatically created profiles**

One way of dealing with alternative word choices and meanings, is the use of a profile-based method, which has become a standard in the field since it was introduced by Geeraerts et al. (1999). We call the 'onomasiological' profile of a given concept the whole of the alternative lexicalisations within a dataset together with the specific frequency distribution. Comparing these profiles between regions, registers and semantic fields gives an overview of the language variation at hand. In a very much similar way, this can also be done with a given word and its alternative meanings, which is then called the 'semasiological' profile of a word. One labour and time-intensive obstacle here is the required manual design and verification of the profile and its content to guarantee a correct picture. Fortunately, strong computational power and large corpora (are on their way to) offer a solution. The idea that the context of a word says something about its meaning can be employed to automatically extract alternative lexicalisations, or synonyms, i.e. words with the same contexts. Similarly, it is possible by looking at the different contexts of a single word to find its meanings. As such, the scale of the investigated profiles can be exponentially enlarged, allowing a much broader and more complete picture of language variation along regional, stratificational and semantic lines.

#### **3.3 Inferential statistics**

Once the data have been gathered, prepared and analysed, inferential statistics allow to draw conclusions regarding the tested hypotheses. From more basic statistical models such as ANOVA to more complex mixed effect linear regression analyses, the most fitting inferential statistics are used to put the data to the test of statistical significance.

### **4 QLVL PhD research projects**

Four of the ongoing PhD projects at the QLVL research group specifically relate to the above-mentioned methods and approaches to the study of Dutch in one way or another. Working on a KU Leuven OT project, Thomas Wielfaert's research focuses on the methodological challenge to automatically disentangle word meanings using distributional modelling. This bottom-up approach requires large text corpora which are stratified for region, register and topic. Visual analytics is used to explore a word's semantic structure and to evaluate the output at the same time. Within the same project, Jocelyne Daems takes a more descriptive point of view and studies onomasiological and semasiological variation in Dutch, and,

more specifically, the link between the two. In line with this profile-based approach, Karlien Franco, funded by an FWO project, makes use of dictionary databases of Dutch dialects to uncover which and how semantic characteristics affect lexical geolinguistic variation. This knowledge can then be employed in dialectometrical analyses. Finally, FWO research fellow Laura Rosseel investigates language attitudes towards colloquial and standard Dutch accents by making use of auditory affective priming. Her basic focus is extended in two methodological directions: on the one hand, she compares AAP to traditional methods in the field. On the other, she tests whether AAP can be used to measure attitudes towards other language levels than just accents (e.g. *words*).

## 5 Conclusion

The QLVL research unit has a strong tradition in quantitative empirical research on language variation in varieties of Dutch. Current PhD projects at QLVL focus predominantly on lexical variation (both from a semasiological and onomasiological perspective and on attitudes towards language and language varieties. The profile-based approach that is employed in several of the PhD projects aims to disentangle the influence of internal and external factors and to provide further insight into the structure of lexical variation in Dutch. Furthermore, by using experimental and corpus data to develop and evaluate advanced methodological tools, the research that is currently being conducted at QLVL also complements and further solidifies the research group's strong position in the field of Cognitive Linguistics.

## References

- Clyne, Michael (ed.). 1992. *Pluricentric languages: Differing norms in Different nations*. Berlin/New York: Mouton de Gruyter.
- Degner, Juliane. 2011. Affective priming with auditory speech stimuli. *Language and Cognitive Processes* 26 (10): 1710–1735.
- Garrett, Peter. 2010. *Attitudes to Language*. Cambridge: Cambridge University Press.
- Geeraerts, Dirk. 2003. Cultural models of linguistic standardization. In: Dirven, Roslyn and Pütz (eds.), *Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings*, 25–68. Berlin: Mouton de Gruyter.
- Geeraerts, Dirk. 2011. Colloquial Belgian Dutch. In: Soares da Silva, Torres and Gonçalves (eds.), *Línguas Pluricêntricas. Variação Linguística e Dimensões Sociocognitivas*, 61–74. Braga: Publicações da Faculdade de Filosofia, Universidade Católica Portuguesa.
- Geeraerts, Dirk and Dirk Speelman. 2010. Heterodox concept features and onomasiological heterogeneity in dialects. In: Geeraerts, Kristiansen and Peirsman (eds.), *Advances in Cognitive Sociolinguistics*, 23–40. Berlin/New York: De Gruyter Mouton.
- Geeraerts, Dirk and Hans Van de Velde. 2013. Supra-regional characteristics of colloquial Dutch. In: Hinsken and Taeldeman (eds.), *Language and Space: An International Handbook of Linguistic Variation. Volume 3: Dutch*. Berlin: De Gruyter Mouton.
- Geeraerts, Dirk, Stefan Grondelaers and Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P.J. Meertens-Instituut.
- Ruette, Tom, Dirk Speelman and Dirk Geeraerts. 2011. Measuring the lexical distance between registers in national varieties of Dutch. In: Soares da Silva, Torres and Gonçalves (eds.), *Línguas Pluricêntricas. Variação Linguística e Dimensões Sociocognitivas*, 541–554. Braga: Publicações da Faculdade de Filosofia, Universidade Católica Portuguesa.
- Schütze, Hinrich. 1998. Automatic word sense discrimination. *Computational Linguistics* 24 (1): 97–123.

- Speelman, Dirk and Dirk Geeraerts. 2009. The role of concept characteristics in lexical dialectometry. *International Journal of Humanities and Arts Computing* 2: 221–242.
- Speelman, Dirk, Adriaan Spruyt, Leen Impe and Dirk Geeraerts. 2013. Language attitudes revisited: Auditory affective priming. *Journal of Pragmatics* 52: 83–92.
- Turney, Peter D. and Patrick Pantel. 2010. From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence* 37 (1):141–188.